



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|-----------------|-------------|-------------------------|---------------------|------------------|
| 09/683,263 | 12/05/2001 | Roger Burrowes Bradford | SAIC0030 | 1023 |

28694 7590 10/06/2003

HOWREY SIMON ARNOLD & WHITE LLP
1299 PENNSYLVANIA AVE., NW
BOX 34
WASHINGTON, DC 20004

EXAMINER

ABEL JALIL, NEVEEN

ART UNIT PAPER NUMBER

2175

DATE MAILED: 10/06/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/683,263

Applicant(s)

BRADFORD ET AL.

Examiner

Neveen Abel-Jalil

Art Unit

2175

-- The MAILING DATE of this communication appears on the reverse with the corresponding address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☐ Responsive to communication(s) filed on ____.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-24 is/are pending in the application.
- 4a) Of the above claim(s) ____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) ____ is/are allowed.
- 6) ☒ Claim(s) 1-24 is/are rejected.
- 7) ☐ Claim(s) ____ is/are objected to.
- 8) ☐ Claim(s) ____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on ____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) ☐ The proposed drawing correction filed on ____ is: a) ☐ approved b) ☐ disapproved by the Examiner.
- If approved, corrected drawings are required in reply to this Office action.
- 12) ☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

- 13) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. ____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.
- 14) ☒ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
- a) ☐ The translation of the foreign language provisional application has been received.
- 15) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121. **DOV POPOVICI**

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) 2.

- 4) ☐ Interview Summary (PTO-413) Paper No(s) ____.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: ____.

SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100

DETAILED ACTION

Specification

1. The title of the invention is not descriptive. A new title is required that is clearly indicative of the invention to which the claims are directed. The following title is suggested:

Method for document comparison and selection using latent semantic content and term tuple.

2. The arrangement of the disclosed application does not conform with 37 CFR 1.77(b). Section headings should not be underlined and/or **boldfaced**. Appropriate corrections are required according to the guidelines provided below:

3. The following guidelines illustrate the preferred layout for the specification of a utility application. These guidelines are suggested for the applicant's use.

Arrangement of the Specification

As provided in 37 CFR 1.77(b), the specification of a utility application should include the following sections in order. Each of the lettered items should appear in upper case, without underlining or bold type, as a section heading. If no text follows the section heading, the phrase "Not Applicable" should follow the section heading:

- (a) TITLE OF THE INVENTION.
- (b) CROSS-REFERENCE TO RELATED APPLICATIONS.
- (c) STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT.
- (d) INCORPORATION-BY-REFERENCE OF MATERIAL SUBMITTED ON A COMPACT DISC (See 37 CFR 1.52(e)(5) and MPEP 608.05. Computer program listings (37 CFR 1.96(c)), "Sequence Listings" (37 CFR 1.821(c)), and tables having more than 50 pages of text are permitted to be submitted on compact discs.) or
REFERENCE TO A "MICROFICHE APPENDIX" (See MPEP § 608.05(a). "Microfiche Appendices" were accepted by the Office until March 1, 2001.)
- (e) BACKGROUND OF THE INVENTION.

Art Unit: 2175

- (1) Field of the Invention.
- (2) Description of Related Art including information disclosed under 37 CFR 1.97 and 1.98.
- (f) BRIEF SUMMARY OF THE INVENTION.
- (g) BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S).
- (h) DETAILED DESCRIPTION OF THE INVENTION.
- (i) CLAIM OR CLAIMS (commencing on a separate sheet).
- (j) ABSTRACT OF THE DISCLOSURE (commencing on a separate sheet).
- (k) SEQUENCE LISTING (See MPEP § 2424 and 37 CFR 1.821-1.825. A "Sequence Listing" is required on paper if the application discloses a nucleotide or amino acid sequence as defined in 37 CFR 1.821(a) and if the required "Sequence Listing" is not submitted as an electronic document on compact disc).

Claim Objections

4. Claims 14 and 18 are objected to because of the following informalities:

In claim 14, line 4; the recitation "_Ref532030037" must be deleted. In claim 18, line 4; the recitation "_Ref532038902" must be deleted. Appropriate correction is required.

Claim Rejections - 35 USC § 102

5. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

6. Claims 1-24 are rejected under 35 U.S.C. 102(e) as being anticipated by Li (U.S. Pub. No. 2002/0059161 A1).

Art Unit: 2175

As to claim 1, Li discloses a method for representing the latent semantic content of a plurality of documents, each document containing a plurality of terms (See page 2, paragraph 0016), the method comprising:

deriving at least one n-tuple term from the plurality of terms (See page 3, paragraph 0035);

forming a two-dimensional matrix, each matrix column c corresponding to a document (See figure 2 (b)),

each matrix row r corresponding to a term occurring in at least one document corresponding to a matrix column (See figures 2(a), and 2(b), wherein “term” reads on “word list”),

each matrix element (r, c) related to the number of occurrences of the term (See page 8, paragraph 0096)

corresponding to the row r in the document corresponding to column c , at least one matrix element related to the number of occurrences of one at least one n-tuple term occurring in the at least one document (See page 8, paragraph 0103, also see page 3, paragraph 35), and

performing singular value decomposition and dimensionality reduction on the matrix to form a latent semantic indexed vector space (See page 2, paragraphs 0015-0016).

As to claim 2, Li discloses comprising:

identifying an occurrence threshold (See page 11, claim 63 language);

wherein n-tuples that appear less times in the document collection than the occurrence threshold are not included as elements of the matrix (See page 7, paragraph 0089, wherein “not included” reads on “filtering out”, also see page 3, paragraph 0035).

As to claim 3, Li discloses wherein the occurrence threshold is two (See page 9, claim 10 language, also see page 10, claim 25 language).

As to claim 4, Li discloses wherein deriving at least one n-tuple term further comprises: creating the at least one n-tuple term from n consecutive verbatim terms (See page 6, paragraphs 0074-0080, also see page 3, paragraph 0035).

As to claim 5, Li discloses a method for determining conceptual similarity between a subject document and at least one of a plurality of reference documents, each document containing a plurality of terms (See page 5, paragraph 0060, wherein “determining conceptual similarity” reads on “grouping dictionary words into semantic concepts”), the method comprising:

deriving at least one n-tuple term from the plurality of terms (See page 3, paragraph 0035),

forming a plurality of two-dimensional matrices wherein (See page 2, paragraph 0015), for each matrix:

each matrix column c corresponds to a document, one column corresponding to the subject document (See figures 2(a), and 2(b), also see page 8, paragraph 0097);

each matrix row r corresponds to a term occurring in at least one document corresponding to a matrix column (See figures 2(a), and 2(b), wherein “term” reads on “word list”),

each matrix element (r, c) represents the number of occurrences of the term corresponding to r in the document corresponding to c (See page 8, paragraph 0096);

performing singular value decomposition and dimensionality reduction on a plurality of formed matrices, to form a plurality of latent semantic indexed vector spaces (See page 2, paragraph 0016, also see page 7, paragraph 0081),

the latent semantic indexed vector spaces including at least one space formed from a matrix including at least one element corresponding to the number of occurrences of at least one n -tuple term in at least one document (See page 2, paragraph 0016, also see page 3, paragraph 0035),

determining at least one composite similarity measure between the subject document and at least one reference document as a function of a weighted similarity measure of the subject document to the reference document in each of a plurality of indexed vector spaces (See page 5, paragraphs 0057-0060, also see page 3, paragraph 0038).

As to claim 6, Li discloses wherein the similarity measures from vector spaces comprising greater numbers of n -tuples are weighted greater than similarity measures from vector spaces comprising lesser number of n -tuples (See page 3, paragraph 0035, also see page 2, paragraphs 0015-0017)

As to claim 7, Li discloses a method for representing a query document, the query document containing verbatim terms, the query document intended for querying a collection of reference documents via a latent semantic indexed representation of the reference collection (See page 2, paragraphs 0016-0017); the method comprising:

identifying verbatim entities (See page figure 8, 1st, shows “verbatim entities” represented by “exact match”);

stemming identified entities;

generalizing stemmed entities (See page 5, paragraph 0057, wherein “stemming” reads on “grouping words by using word stemming”); and

supplementing verbatim entities with corresponding generalized entities (See page 2, paragraph 0018).

As to claim 8, Li discloses a method for representing a query document, the query document containing verbatim terms, the query document intended for querying a collection of reference documents via a latent semantic indexed representation of the reference collection (See page 8, paragraph 0104, also see page 2, paragraphs 0016-0017); the method comprising:

identifying verbatim entities (See page figure 8, 1st, shows “verbatim entities” represented by “exact match”);

stemming identified entities;

generalizing stemmed entities (See page 5, paragraph 0057, wherein “stemming” reads on “grouping words by using word stemming”); and

Art Unit: 2175

replacing verbatim entities with corresponding generalized entities (See page 4, paragraph 0040, also see page 4, paragraphs 0043-0047).

As to claim 9, Li discloses wherein verbatim entities comprise ordered terms between stop words (See page 5, paragraph 0057).

As to claim 10, Li discloses wherein generalizing entities further comprises alphabetically ordering stemmed entities as an aid to generalization (See pages 7-8, paragraphs 0092-0095).

As to claim 11, Li discloses wherein generalizing entities further comprises ordering stemmed entities as a function of the frequency of occurrence of verbatim entities (See page 5, paragraph 0057-0060).

As to claim 12, Li discloses wherein generalized entities are identified with human feedback (See page 2, paragraph 0018).

As to claim 13, Li discloses wherein generalized entities are identified by automated process (See page 5, paragraph 0056, wherein “automated process” reads on “automatic view”).

As to claim 14, Li discloses a method for characterizing the results of a query into a latent-semantic indexed document space, the query comprising at least one term, the results

Art Unit: 2175

comprising a set of document identities (See page 2, paragraph 0016, also see page 4, paragraph 0040); the method comprising:

ranking results as a function of at least the frequency of occurrence of at least one term (See page 3, paragraph 0035, also see page 7, paragraph 0092).

As to claim 15, Li discloses wherein at least one term used in ranking is a query term (See page 3, paragraph 0035, also see page 7, paragraph 0092).

As to claim 16, Li discloses wherein the at least one query term used in ranking is a generalized entity (See page 3, paragraph 0035, also see page 7, paragraph 0092).

As to claim 17, Li discloses wherein the at least one term used in ranking is a generalized entity (See page 3, paragraph 0035, also see page 7, paragraph 0092).

As to claim 18, Li discloses a method for determining conceptual similarity between a query document and at least one of a plurality of reference documents, each document comprising a plurality of verbatim terms, the reference documents indexed into a latent semantic index space (See page 2, paragraph 0016, also see page 4, paragraph 0040), the method comprising:

identifying verbatim entities (See page figure 8, 1st. shows “verbatim entities” represented by “exact match”);

stemming identified entities;

generalizing stemmed entities (See page 5, paragraph 0057, wherein “stemming” reads on “grouping words by using word stemming”);

replacing at least one verbatim entity with the corresponding generalized entity to form a generalized query (See page 4, paragraph 0040, also see page 4, paragraph 0048);

identifying a set of reference documents based on closeness, within the latent semantic indexed space, between the generalized query and each reference document (See page 4, paragraphs 0038-0040); and

ranking a subset of closest identified documents as a function of at least the frequency of occurrence of at least one term (See page 5, paragraphs 0060-0062, also see page 6, paragraph 0071).

As to claim 19, Li discloses wherein at least one term used in ranking is a query term (See page 3, paragraph 0020).

As to claim 20, Li discloses wherein the at least one query term used in ranking is a generalized entity (See page 3, paragraph 0020).

As to claim 21, Li discloses wherein the at least one term used in ranking is a generalized entity (See page 3, paragraph 0020).

Art Unit: 2175

As to claim 22, Li discloses a method for representing the latent semantic content of a plurality of documents, each document containing a plurality of verbatim terms (See page 2, paragraphs 0016-0017), the method comprising:

deriving at least one expansion phrase from the verbatim terms, each expansion phrase comprising terms (See page 2, paragraph 0017, also see page 6, paragraph 0072);

replacing at least one occurrence of a verbatim term having an expansion phrase with the expansion phrase corresponding to that verbatim term (See page 3, paragraph 0020, also see page 8, paragraph 0092);

forming a two-dimensional matrix (See page 8, paragraph 0096, also see page 8, paragraph 0097),

each matrix column c corresponding to a document (See figure 2(b));

each matrix row r corresponding to a term (See page 2, paragraph 0015);

each matrix element (r, c) representing the number of occurrences of the term corresponding to r in the document corresponding to c (See page 8, paragraph 0096);

at least one matrix element corresponding to the number of occurrences of one at least one term occurring in the at least one expansion phrase (See page 8, paragraphs 0097-0099), and

performing singular value decomposition and dimensionality reduction on the matrix to form a latent semantic indexed vector space (See page 2, paragraph 0016, also see page 7, paragraph 0081).

As to claim 23, Li discloses a method for representing the latent semantic content of a plurality of documents, each document containing a plurality of terms, the method comprising:

identifying at least one idiom among the documents (See page 3, paragraph 0035, wherein “idiom” reads on “syntactically”),

each idiom containing at least one idiom term (See page 7, paragraph 92, wherein “idiom” reads on “syntactically”);

forming a two-dimensional matrix (See page 8, paragraph 0096, also see page 8, paragraph 0097),

each matrix column corresponding to a document (See figure 2(b));

each matrix row corresponding to a term occurring in at least one document represented by a row (See figure 2(a));

each matrix element representing the number of occurrences of the term corresponding to the element's row in the document corresponding to element's column (See page 8, paragraph 0096);

at least one occurrence of at least one idiom term being excluded from the number of occurrences corresponding to that term in the matrix (See page 7, paragraph 0089, wherein “excluded” reads on “filtering out”, also see page 3, paragraph 0035),

performing singular value decomposition and dimensionality reduction on the matrix (See page 2, paragraph 0016, also see page 7, paragraph 0081).

As to claim 24, Li discloses a method for representing the latent semantic content of a plurality of documents, each document containing a plurality of terms, the method comprising:

identifying at least one idiom among the documents (See page 3, paragraph 0035, wherein “idiom” reads on “syntactically”),

each idiom containing at least one idiom term (See page 7, paragraph 92, wherein “idiom” reads on “syntactically”);

replacing at least one identified idiom with a corresponding idiom elaboration, each elaboration comprising at least one elaboration term (See page 9, claim 13 language),

forming a two-dimensional matrix (See page 8, paragraph 0096, also see page 8, paragraph 0097),

each matrix column corresponding to a document (See figure 2(b));

each matrix row corresponding to a term (See figure 2(a));

each matrix element representing the number of occurrences of the term corresponding to the element's row in the document corresponding to element's column (See page 8, paragraph 0096),

at least one matrix element corresponding to the number of occurrences of an elaboration term in a document corresponding to a matrix column (See page 8, paragraphs 0096-0097);

performing singular value decomposition and dimensionality reduction on the matrix (See page 2, paragraph 0016, also see page 7, paragraph 0081).

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Paik et al. (U.S. Patent No. 6,263,335 B1) teaches information extraction system and method using concept-relations-concept (CRC) triples.

Chidlovskii (U.S. Patent No. 6,347,314 B1) teaches answering queries using query signatures and signatures of cached semantic regions.

Aggarwal et al. (U.S. Patent No. 6,349,309 B1) teaches system and method for detecting clusters of information with application to e-commerce.

Hofmann et al. (U.S. Pub. No. 2002/0107853 A1) teaches system and method for personalized search, information filtering, and for generating recommendations utilizing statistical latent class models.

Anick et al. (U.S. Patent No. 6,519,586 B2) teaches method and apparatus for automatic construction of faceted terminological feedback for document retrieval.

Berghofer et al. (U.S. Pub. No. 2003/0088480 A1) teaches enabling recommendation systems to include general properties in the recommendation process.

8. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Neveen Abel-Jalil whose telephone number is 703-305-8114. The examiner can normally be reached on 8:00AM-4: 30PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dov Popovici can be reached on 703-305-3830. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

Neveen Abel-Jalil

Application/Control Number: 09/683,263

Page 15

Art Unit: 2175

September 17, 2003



DOV POPOVICI
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100